DOCUMENT RESUME

ED 090 279 TM 003 558

AUTHOR Kulm, George
TITLE The Effects of the Two Summative Evaluation Methods on Achievement and Attitudes in Individualized Seventh-Grade Mathematics.
PUB DATE [73]
NOTE 18p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Ability Grouping; *Achievement Gains; Grade 7; Grouping (Instructional Purposes); *Individualized Instruction; *Mathematics; Student Ability; Student Attitudes; Student Testing; *Summative Evaluation

ABSTRACT

One hundred fifty-nine seventh grade mathematics students were divided into four treatment groups: 1) traditional, tested after each unit, 2) individualized, tested after each unit, 3) individualized, tested after each objective, and 4) individualized, chose to be tested after each objective. Each group was divided into high and low math ability and the achievement and attitude data were analyzed using analyses of variance. No achievement differences were found, but significant interactions indicated that low ability students with a choice of testing methods, individualized instruction, and tested after each unit had better attitudes toward math than other groups of low ability students. High ability students preferred objective testing and were satisfied with the status quo. (Author)

ED 090279

28.06

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Effects of Two Summative Evaluation Methods on Achievement and Attitudes in Individualized Seventh-Grade Mathematics

Gerald Kulm

Purdue University

T.. 003 558

The Effects of Two Summative Evaluation Methods
on Achievement and Attitudes in
Individualized Seventh-Grade Mathematics

ABSTRACT

159 seventh grade mathematics students were divided into four treatment groups: T (traditional, tested after each unit), I1 (individualized, tested after each unit), I2 (individualized, tested after each objective), and I3 (individualized, chose to be tested after each objective). Each group was divided into high and low math ability and the achievement and attitude data were analyzed by 2×2 ANOVAS. No achievement differences were found, but significant interactions indicated that low ability students with a choice of testing methods, individualized instruction, and tested after each unit had better attitudes toward math than other groups of low ability students. High ability students preferred objective testing and were satisfied with the status quo.

The Effects of Two Summative Evaluation Methods on Achievement and Attitudes in Individualized Seventh-Grade Mathematics

The opportunity for individual decision-making is a key aspect of individualized instruction. One of the choices most often available to a student is the rate of progress through the material. There is evidence that self-pacing may not accomodate individual differences in ability. Newmark (1970) found that work rate of 8th graders in mathematics was independent of cognitive variables. On the other hand, Yeager and Lindvall (1967) reported that rate of progress was correlated with achievement in mathematics but not in reading. Wang (1968) also found rate of learning to be task specific and not a general factor that characterized student achievement in all learning situations. These results indicate that individualized instruction in which self-pacing is the only choice available may not be more effective than more traditional approaches.

The method of evaluation is also an important component of an individualized approach. Pretests, self-tests, and formative tests are often used to diagnose and monitor progress and prescribe instruction. Mastery learning theory (Bloom 1968) suggests that formative (non-graded) tests over small units should be used to provide students with feedback on their progress. Summative (graded) tests should be given after several units have been mastered. Since students are not graded during learning, this procedure

should produce improved attitudes and self-concepts. Research results have not always indicated that the mastery approach to evaluation, adopted in many individualized systems, has the predicted effects. Block (1971) found that maintenance of a high level of mastery on formative tests produced negative effects on the attitudes and interests of 8th graders studying algebra. His study lasted only a week, and it may be that over longer period this effect would be even more pronounced. In a study of mastery learning in college freshman mathematics, Wheatley, et al (1973) did not find improved achievement or attitudes using formative evaluation.

The generally accepted theory that formative evaluation is less threatening may not account for the anxiety that students feel when they are tested, for a grading purpose, over large amounts of material. Many students seem to prefer frequent quizzes on which to base their grade.

The following hypotheses, in null form, were proposed for the present study.

(1) No achievement or attitude differences will be found between individualized and group instructed students.

(2) No achievement or attitude differences will be found between students who choose or do not choose their method of summative evaluation.

(3) No achievement or attitude differences will be found between students evaluated after each objective and

students evaluated after each unit.

METHOD

Subjects: The subjects were 159 seventh grade mathematics students in an Indiana junior high school.

Treatments: At the beginning of the second semester, all of the seventh grade class was divided into two groups, "traditional" and "individualized." The placement of students was based on teachers' judgments of the probable success of students in an individualized setting. After the first week, a few transfers were made, then all students remained in the groups for the 6 week period of the study.

The traditional (T) groups (2 classes of 25 Ss per period) were taught by teachers who preferred that method. The individualized groups (about 50 Ss per period) were taught by a team leader, an intern, and a paraprofessional. At the beginning of the semester, the individualized group that met during 1st period (I1) was told that they would be tested at the end of each unit (about 6-8 objectives) of work. The 2nd period group (I2) was told that they would be tested at the end of each objective. The 3rd period group (I3) was told that they could choose between the two testing methods. The choice was nearly unanimous for testing after each objective, so the entire group decided to use that method.

All groups studied geometry using the same teacher-prepared materials, consisting of a single lesson written for each objective. The individualized groups took a self-test at the end of each lesson, then either studied further or (1) took a summative test (I2 and I3) or (2) began the next objective, repeating until taking a unit test (I1). In groups I2 and I3, the different forms of the summative test were retaken until a score of 70% was obtained. The traditional groups were tested at the end of the same units as I1, but had teacher-directed review rather than self-tests.

Design: Since the traditional groups were evaluated only by the unit test method, and group I3 chose the single objective method, it was necessary to use an incomplete factorial design. In addition, ability interactions with the treatments were thought to be possible, so the groups were divided into high and low mathematics ability by using the median score of the 159 Ss on the Iowa Composite Arithmetic test. The groups compared may be illustrated in the diagram below.

	Unit		Objective	
	Teaching Method		Choice of Method	
	Trad	Indiv	No	Yes
Ability High	T	I1	I2	I3
Low	T	I1	I2	I3

Tests: At the end of the 6-week period, all students were given the same teacher-made achievement test over the geometry objectives. In addition, an attitude test made up of 20 items selected from the NLSMA Battery (Wilson, et al 1968) was administered. The students were also asked 4 questions about their reaction to the method of testing used in their group.

1. During the past six weeks do you think that the method of testing you used showed how much you really knew?

   a) no b) sometimes not c) sometimes yes d) yes

2. How did the testing method during the past six weeks compare with the way you have usually been tested in math?

   a) not like before b) a little like before
   c) quite a bit like before d) just like before

3. Would you like to have the same testing method again for the next six weeks?

   a) definitely not b) maybe not c) maybe yes
   d) definitely yes

4. Did the testing method during the last six weeks help you learn math better than other ways you have been tested?

   a) a lot less b) a little less c) a little more
   d) a lot more

RESULTS

The mean achievement and attitude scores for each treatment group are given in Table 1. Traditional is denoted T, high ability, H; and low ability, L. The attitude subscale Math Fun vs Dull was derived from 3 items of the same name in the NLSMA Battery. Some of the students misunderstood directions and did not write their names on the attitude response card and some answers were omitted, so the number of subjects for each variable is different.

Insert Table 1 about here

Three separate 2 × 2 ANOVAs were used to compare the groups in the incomplete factorial design. Table 2 summarizes the Ability by Test Method analysis.

Insert Table 2 about here

The significant interaction on the Fun vs Dull attitude scale indicated that low ability students evaluated by units thought math was fun, but high ability students evaluated on each objective thought math was fun. Although significant at only the .09 level, the interaction on the total attitude score had this same pattern; low ability Ss having better attitudes under unit testing while high ability Ss had better attitudes when tested after each objective.

The questionaire results indicated that students tested after each unit thought the testing method showed what they really knew. Also, high ability students viewed

the testing method more like previous methods than did low ability students.

The Ability by Choice of Test Method analysis is summarized in Table 3.

Insert Table 3 about here

The significant interaction on the Fun vs Dull Attitude scale revealed that low ability students who had a choice of methods felt math was fun whereas high ability students who were assigned a test method thought math was fun.

The significant effects for ability on questionaire items 2, 3, and 4 favored high ability students. High ability students thought the testing method was like before, they wanted to use the method again, and they felt the method helped them learn better.

The final analysis, Ability by Teaching Method is given in Table 4.

Insert Table 4 about here

The Fun vs Dull attitude scale was again sensitive in detecting differences between groups. The individualized group thought math was more fun than did the traditional group. The interaction showed that while high ability students did not differ much in their attitudes in the two groups, low ability students in the individualized groups thought math was more fun than did low ability students in the traditional groups. This same pattern held for the

teaching method effect and the interaction on the total attitude test.

The questionaire results showed that the individualized groups felt the unit-testing method showed what they really knew and they wanted to use the method again. The significant interaction on question 3 revealed that while high ability students in both groups were satisfied, low ability students in the traditional groups did not want the same testing method again but low ability students in the individualized groups definitely wanted to use the unit-testing method again.

DISCUSSION

A number of patterns appeared to emerge in the analysis of the achievement and attitude data. First, none of the null hypotheses concerning achievement were rejected. The individualized teaching method, the method of summative evaluation, and the opportunity to choose the method of evaluation all failed to produce differences in achievement. One could speculate at length about the reasons or explanations for this result, but apparently, the various situations simply did not differ enough from each other to produce differences in learning. Students seem to spend a certain amount of time or energy in learning mathematics and it appears that more comprehensive changes in the instructional setting than those in this study are needed to increase the students' willingness to spend more time and effort.

The null hypotheses concerning attitudes were not supported. The treatments in the study had significant effects on attitudes, particularly for lower ability students. It appeared that the higher ability students were satisfied with things, no matter what was done. They were not anxious for change and felt that their achievement was measured accurately. On the other hand, the low ability students reacted positively to having a choice of testing methods and to learning in an individualized setting. They perceived changes in testing procedures and were not content with doing the same things again.

These results indicate quite clearly that traditional teaching methods may be inappropriate for low ability students. The frustration of failure was not accepted as a fact of life by these students. The results revealed that low ability students reacted positively to changes in the learning environment. If it was possible to improve attitudes, perhaps achievement could be improved over a greater period of time.

The low ability students thought math was more fun when unit-tested. Perhaps the continuous testing in the objective method was frustrating to them, especially if they had to repeat the test. High ability students enjoyed this continuous positive reinforcement. This result may indicate that while it is desirable to present small bits of content to low ability students, one should be careful not to frustrate the student by continual evaluation.

Because this study was school-based and imperfectly controlled, it has some obvious limitations. The data was incomplete because of absences and other factors, making the number of complete data cases in some of the experimental groups quite low.

Although students in the individualized treatment worked at their own pace, they were required to complete the objectives by the end of the 6-week period. This requirement may have negated some of the possibilities for mastery by lower ability students.

Finally, although adjustments were made, the teachers' judgments of students likely to succeed in an individualized setting was clearly not a random assignment to groups. Better attitudes in the individualized group may have been due partially to this special treatment effect. On the other hand, if this select group did not achieve better, it appears that in a random assignment to teaching methods, the individualized approach might have fared worse. This method of selection would not seem to negate, however, the results related to the attitudes of low ability students. It seems apparent that these are the students who need more individual attention and freedom of choice in the learning environment.

REFERENCES

Block, J. H. The effects of various levels of performance on selected cognitive, affective, and time variables. Summary of doctoral dissertation in Block, J. H. (ed.) Mastery learning, New York: Holt, Rinehart and Winston, 1971.

Bloom, B. S. Learning for mastery. UCLA-CSEIP Evaluation comment, 1, No. 2, 1968.

Newmark, G. The relationship between student characteristics and work rate and between work rate and performance in programmed instruction with two different subject matter fields. Dissertation Abstracts, 1970, 31, 1145.

Wang, M. An investigation of selected procedures for measuring and predicting rate of learning in classrooms operating under a program of individualized instruction. Dissertation Abstracts, 1968, 29, 1176.

Wheatley, G. W., Kulm, G. and Kane, R. B. Objectives and formative evaluation as factors in mathematics achievement and attitudes of college students in a number systems course. Paper presented at NCTM Annual Meeting, Houston, 1973.

Wilson, J. W., Cahen, L. S. and Begle, E. G. Y-population test batteries, NLSMA Reports, No. 2, Part B. Stanford University, 1968.

Yeager, J. L. and Lindvall, C. M. An exploratory investigation of selected measures of rate of learning. Journal of Experimental Education, 1967, 36, 78-81.

Table 1

Means and Standard Deviations of Dependent Variables for Treatment Groups

Treatment Group[a]	Dependent Variable: 6-week Test			Math Fun vs Dull			Total Attitude			Question 1			Question 2			Question 3			Question 4		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD
[illegible] L	47	32.9	11.9	39	7.5	2.2	39	39.7	7.2	38	3.0	1.1	23	2.6	.8	22	2.1	.9	22	2.6	.8
[illegible] H	13	43.7	7.7	11	8.8	2.1	11	44.3	6.4	11	2.9	1.0	9	2.3	1.2	9	3.0	.9	7	2.8	1.1
[illegible] L	14	32.7	11.7	9	10.3	1.4	9	48.0	6.4	9	3.5	.5	9	2.5	.9	9	3.4	.7	8	3.1	.8
[illegible] H	19	47.2	4.5	10	8.6	1.9	10	43.5	7.7	10	3.5	.7	10	3.0	.8	10	2.8	1.1	10	2.7	.9
[illegible] L	12	35.8	10.7	5	6.8	4.1	5	41.6	11.5	5	2.6	1.5	5	2.2	1.3	5	2.8	1.6	5	3.0	1.2
[illegible] H	22	47.1	5.8	15	9.8	1.7	15	46.3	7.4	15	3.0	1.1	15	3.4	.5	14	3.1	1.0	14	3.3	.7
[illegible] L	6	37.7	10.0	3	8.7	3.1	3	41.3	3.5	2	2.5	2.1	2	3.0	0.0	2	1.5	.7	2	1.5	.7
[illegible] H	26	46.4	6.1	14	7.3	1.6	14	39.5	6.8	14	3.4	1.1	13	3.3	.7	13	3.3	1.1	13	3.3	.9
Total	159	40.1	11.0	106	8.3	2.3	106	42.3	7.6	104	3.1	1.1	86	2.3	1.4	84	2.2	1.5	81	2.2	1.5

[a] L denotes low math ability, H denotes high math ability

Table 2

ANOVA of Achievement and Attitude Scores for Individualized Groups (Ability × Test Method)

Source	df	MS	F	p
		Six-Week Test		
Ability (A)	1	2624.36	40.11	.0001
Test Method (T)	1	33.76	0.52	.52
A × T	1	39.52	0.60	.55
Error	63	65.43		
		Math Fun vs Dull		
Ability	1	3.36	0.74	.60
Test Method	1	11.39	2.49	.12
A × T	1	46.89	10.28	.003
Error	35	4.56		
		Total Attitude		
Ability	1	.11	0.01	.96
Test Method	1	26.63	0.43	.52
A × T	1	178.44	2.87	.09
Error	35	62.09		
		Question 1		
Ability	1	.25	.26	.62
Test Method	1	4.43	4.57	.04
A × T	1	.43	.45	.51
Error	35	.98		
		Question 2		
Ability	1	5.66	8.76	.005
Test Method	1	.01	.01	.93
A × T	1	1.19	1.85	.18
Error	35			
		Question 3		
Ability	1	.19	.16	.69
Test Method	1	.19	.16	.69
A × T	1	2.02	1.70	.20
Error	35	1.19		
		Question 4		
Ability	1	.01	.01	.91
Test Method	1	.57	.72	.59
A × T	1	1.23	1.55	.22
Error	33	.79		

Table 3

ANOVA of Achievement and Attitude Scores

for Individualized Groups (Ability × Choice of Test Method)

Source	df	MS	F	p
		Six-Week Test		
Ability (A)	1	1199.97	21.73	.0001
Choice (c)	1	4.09	.07	.78
A × C	1	19.95	.36	.56
Error	62	55.21		
		Math Fun vs Dull		
Ability	1	4.26	.87	.64
Choice	1	.49	.10	.75
A × C	1	27.66	5.67	.02
Error	33	4.88		
		Total Attitude		
Ability	1	12.53	.21	.65
Choice	1	75.08	1.28	.26
A × C	1	64.22	1.09	.30
Error	33	58.63		
		Question 1		
Ability	1	2.11	1.43	.24
Choice	1	.13	.09	.77
A × C	1	.33	.23	.64
Error	32	1.47		
		Question 2		
Ability	1	2.69	4.86	.03
Choice	1	.59	1.07	.31
A × C	1	.94	1.70	.20
Error	31	.55		
		Question 3		
Ability	1	5.45	4.11	.049
Choice	1	1.52	1.14	.29
A × C	1	2.53	1.91	.17
Error	30	1.33		
		Question 4		
Ability	1	5.52	6.77	.014
Choice	1	2.83	3.47	.07
A × C	1	2.48	3.04	.09
Error	30	.82		

Table 4

ANOVA of Achievement and Attitude Scores for Unit-tested Groups (Ability × Teaching Method)

Source	df	MS	F	p
		Six-week Test		
Ability	1	2864.88	27.31	.0001
Method	1	48.18	.46	.51
A × M	1	61.17	.58	.55
Error	89	104.90		
		Math Fun vs Dull		
Ability	1	.63	.14	.71
Method	1	20.26	4.58	.03
A × M	1	27.71	6.26	.014
Error	65	4.43		
		Total Attitude		
Ability	1	.02	.01	.98
Method	1	173.28	3.47	.06
A × M	1	251.64	5.04	.026
Error	65	49.95		
		Question 1		
Ability	1	.06	.07	.79
Method	1	4.00	4.29	.04
A × M	1	.01	.01	.95
Error	64	.93		
		Question 2		
Ability	1	.12	.14	.71
Method	1	1.18	1.39	.24
A × M	1	1.25	1.47	.23
Error	47	.85		
		Question 3		
Ability	1	.13	.16	.70
Method	1	3.34	4.00	.049
A × M	1	6.19	7.41	.009
Error	46	.83		
		Question 4		
Ability	1	.06	.07	.78
Method	1	.34	.42	.53
A × M	1	1.16	1.41	.24
Error	43	.82		